

# Academic Paper Knowledge Graph, the Construction and Application

Xinyu Du and Ning Li

*Beijing Information Science & Technology University, Beijing, China*

## Abstract

Academic papers in the form of documents are still the primary carrier of academic publications. Nevertheless, it is difficult for such documents to express the papers' semantic elements and discourse structures directly. Hence, this paper focuses on knowledge units with semantic information for papers to construct a knowledge graph, affording quickly retrieving knowledge from academic papers. Based on the in-depth analysis of the general narrative regulations of academic papers, we develop an academic paper representation ontology PEO that includes 29 classes, 18 relations, and five attributes. The experiment demonstrates that the developed ontology has a strong ability to represent knowledge of academic papers. Additionally, this paper preliminarily constructs the knowledge graph PKG of academic papers based on PEO ontology, demonstrating its role in semantic retrieval and intelligent question answering. Overall, this study enriches the academic knowledge's expression ability and helps better explore the value of academic papers.

## Keywords

academic papers, ontology, semantic description, knowledge representation, knowledge graph

## 1. Introduction

In recent years, knowledge graphs, as a form of structured human knowledge, have attracted significant research attention in academia and industry and have been widely used in AI tasks such as natural language understanding, question answering, and recommendation systems [1]. With the digital transformation of academic work, applying knowledge graphs in knowledge representation, knowledge mining, knowledge retrieval, and other aspects of the academic literature has become a research hotspot. However, most of the early research was limited to constructing knowledge graphs for the external features of academic papers (e.g., title, author, institution, keywords, issues, and publisher), phrases, key terms, and other knowledge content [2-5]. Recently, some scholars have constructed knowledge graphs for the semantic knowledge of academic papers (e.g., background, methods, results, and conclusions), but the semantic knowledge is incomplete, as it does not realize complex semantic retrieval and question answering [6-9]. For example, "Is there any literature mentioning that a certain method is used to solve a problem?", "For a certain goal, what methods have been proposed in the existing research and how effective?", "What is the best experimental result of a method?". However, under the massive literature resources, current knowledge service platforms, e.g., HowNet, Wanfang, and Baidu Academic, provide literature retrieval methods only from the perspective of article title, subject, author, unit, keywords, abstracts, references, Chinese library classification number, and literature sources. Therefore, the retrieval results often provide the whole literature or text, still requiring manual screening by searchers and then carefully reading the screened documents. This strategy does not meet the scientific researchers' needs to acquire knowledge and information accurately and efficiently. Thus, to realize the above-mentioned intelligent question answering and retrieval, we must build a specific knowledge base that contains the semantic knowledge in academic papers, such as questions, methods, results, and conclusions. However, the authors of academic manuscripts typically



linearly express in natural language, and directly obtaining the papers' semantic knowledge is challenging. Therefore, this study investigates how to define a suitable ontology for the knowledge representation of academic papers and how to construct the knowledge graph of papers based on the ontology.

As a knowledge representation method, ontology can also be employed as the skeleton and foundation of a knowledge base, describing text information from the semantics and knowledge aspect. Ontology is widely used in knowledge representation of academic literature, knowledge management semantic retrieval, scientific argument analysis, and other applications [10]. According to the principles of knowledge units, ontology, and knowledge graph, this paper combines subdividing the paper's content, associating knowledge units, and forming a knowledge network through analyzing the content of the academic papers, determining the knowledge types, and defining the ontology concepts. Then the concepts and relationships in the ontology are used to describe the knowledge units contained in the academic papers and the relationships between them. Finally, a structured semantic knowledge base (namely a knowledge graph) is constructed based on ontology to achieve semantic retrieval and intelligent question answering for academic papers.

## **2. Related Work**

### **2.1. Research Status of Academic Knowledge Graph**

Currently, constructing academic knowledge graphs is a hot research topic that has recently been included in the guideline of national key R&D projects. In 2017, Tsinghua University and Microsoft Research [2] jointly released the Open Academic Graph (OAG), which combines metadata from 155 million academic papers in the ArnetMiner Academic Graph [3] and 160 million papers in the Microsoft Academic Graph (MAG) [4]. The employed data types include the paper's title, author, conference, year, and abstract. Subsequently, the OAG 2.0 version released in 2019 added three types of entities: papers, authors, and publication locations and their corresponding matching relationships. OAG integrates a large amount of paper metadata information, provides intelligent services through data sharing, and promotes the development of academic knowledge graphs. Bratsas et al. [5] constructed a scientific knowledge graph by semantically annotating and linking academic research fields, including all research fields in each scientific field in a standard hierarchy. The above research is of great value in improving the literature's retrieval efficiency. In recent years, knowledge graph research for academic papers has shifted from paper metadata to deep semantic knowledge in papers. For instance, Auer et al. [6] proposed the Open Research Knowledge Graph (ORKG), which describes research contributions traditionally described in scientific articles in a structured and semantic manner. Articles are added to ORKG by retrieving (or manually adding) key metadata for articles from CrossRef via DOI and then using dedicated input fields to describe the content of the research articles. The description includes the research questions, the materials and methods used, and the results obtained so that the research contribution is comparable to other articles addressing the same research question. Fathalla et al. [7] proposed the SemSur ontology for describing the content of the literature review, including four core concepts of research questions, methods, implementation, and evaluation. Then, based on the ontology, a review knowledge graph is generated. Cao and Zhao [8] mined innovative content by extracting innovative sentences in papers for entity recognition and building a knowledge graph for innovative content in academic papers. Roa et al. [9] created a Deep Knowledge Graph (DKG) repository for papers related to deep learning algorithms and methods to help improve the search and retrieval of relevant information in the academic field.

The above research shows that the existing research on academic knowledge graphs is limited to the paper's external features and bibliographic information. Only a few scholars researched the graph construction for the intrinsic semantic knowledge of academic papers. However, the semantic knowledge is not comprehensive enough to cope with the complicated semantic retrieval and question-and-answer for academic papers.

## 2.2. Research Status of Knowledge Representation in Academic Papers

Many scholars analyze the semantic description of the literature content from different perspectives and have proposed different ontologies and models. For example, Groza[11] proposed the SALT framework (Semantically Annotated LaTeX) for document semantic annotation. This framework indexed early document rhetorical units, including document ontology, rhetorical ontology, and annotation ontology. The rhetorical ontology is expanded based on the ABCDE model[12], including abstract, motivation, scenario, contribution, evaluation, discussion, background, conclusion, and entity, and also defines 11 rhetorical relations such as antithesis, circumstance, and concession. SALT has a rough definition of component granularity, which can not describe in detail the content information of each part of the academic paper, but its classification system and relationship definition provide a reference for the related research. Liakata et al.[13] developed a core scientific concept (CoreSCs) that reflected the structure and type of knowledge of scientific research. In 2011, W3C (World Wide Web Consortium) [14] released the Ontology of Rhetorical Blocks (ORB), which creates a general coarse-grained collection of rhetoric modules for scientific publications and provides fine-grained semantic entry for document contents and forms. The Pattern Ontology (PO) constructed by Iorio et al. [15] focused on the attribute description of structural components such as sentences, paragraphs, and chapters. Ribaupierre et al. [16] proposed a user-centric scientific literature annotation model—SciAnnotDoc. However, the academic papers' semantic content description in these studies is not detailed and comprehensive, and the granularity is relatively coarse. Therefore, most scholars further built ontologies for the semantic description of academic paper content at a fine-grained level. Shotton et al. [17] proposed the Discourse Elements Ontology (DEO), which draws on some of the rhetorical structural elements of the rhetorical ontology in the SALT framework, defines components with different rhetorical functions such as background, conclusion, and data, and provides a structured vocabulary for the rhetorical elements in documents. DEO can describe the paper's rhetorical units in detail but does not define their relationship. The Document Components Ontology (DoCO) [18][19] provides a structured vocabulary that defines document components such as title, abstract, chapter, sentence, and paragraph. However, it only provides a fine-grained description of the dissertation structure. Qin et al. [20] introduced a knowledge element ontology model for the knowledge representation of scientific literature, which hierarchically represents the contents of papers and defines the apposition and hierarchical relationships. This model describes the internal and external characteristics of scientific literature in fine granularity, playing a significant role in the deep knowledge service of scientific literature. Based on the work of Zhang et al. [22], Wang et al. [21] constructed the Functional Units Ontology, FUU) of scientific papers, which included 12 first-level categories and 28 second-level categories. This ontology builds a fine-grained model of the organizational structure of scientific papers from the perspective of semantic functions of content components. FUU describes the content components of scientific papers in more detail and reveals better the semantic functions of functional units of scientific papers, having a positive significance for the semantic description of academic papers. However, the ontology does not consider the definition of the relationship between the functional units, and thus it can not represent the logical relationship among various functions. Sun et al. [23] constructed the semantic annotation ontology of academic literature based on inheriting the existing annotation ontology (such as DEO, DoCO, C4O[24], FaBiO, and CiTO[25]). Although the annotation ontology involves the types of academic documents, scientific discourses, structural elements, and references, it cannot comprehensively and carefully describe the content semantics of academic documents. Niu and Ou [26] suggested a semantic annotation framework when exploring the semantic annotation model of scientific papers. This framework realizes the semantic annotation function of the paper's physical and argument structure, with the annotation ontology adopting ORB, scientific experiment ontology (EXPO) [27], the micro-publication ontology [28], and the nano-publication ontology [29]. Although it covers the paper's physical and argument structures, this framework lacks some basic semantic units, such as research background, research questions, and future work.

Additionally, some scholars proposed different models or ontologies from the perspective of scientific argumentation to divide the article's content. For example, Teufel [30] proposed an Argumentative Zoning (AZ) model for analyzing the scientific papers' argumentation and rhetorical

structure. Since the annotation experiments of the model are limited to computer linguistics, Teufel et al. [31] extended and updated AZ and obtained the Argumentative Zoning II (AZ-II) model. Soldatova et al. [27] proposed EXPO, while Vitali et al. [32] introduced an Argument Model Ontology (AMO) based on the Toulmin Argument Model. Wang et al. [33] suggested the scientific paper argumentation ontology SAO, which is used to reveal the important viewpoints, conclusions, and demonstration processes of scientific papers. Qu and Ou [34] constructed a sentence-level and entity-level scientific paper argument structure ontology. Scientific argumentation is a critical process in an academic paper, where the argument model or argument ontology considers the necessary elements of scientific argumentation. Although it is impossible to describe the article's content comprehensively, it still has good reference value for the semantic description of the academic papers' content.

Nevertheless, existing research on semantically describing the literature content has the following deficiencies. 1) It is difficult to reveal the document's semantic units in a detailed and comprehensive manner by simply using rhetorical elements such as methods, results, and conclusions to describe the document's content in coarse-grained semantics. 2) Defining the relationship between semantic units or relying on a trivial definition to reflect the logical relationship between the semantic units of academic papers.

Spurred by the above deficiencies, this paper develops an academic paper representation ontology (PEO) based on the current results to express the semantic units in academic papers in a detailed and comprehensive manner. Moreover, our model provides a basis for constructing academic papers knowledge graphs and realizes academic Semantic retrieval of resources and intelligent question answering.

### **3. Construction of Academic Paper Expression Ontology**

Based on the existing literature relevant to content representation ontology and modeling, this paper determines the types of semantic units through semantic annotation and analysis appropriate for academic paper content. Then, it draws on the argumentation relationship in argumentation structure[33], the rhetorical relationship in rhetorical structure[35], the discourse relations in discourse analysis [36], and the relations defined in existing ontologies, and finally determined 29 classes, 18 relations, and five attributes.

#### **3.1. Class Design in PEO**

This paper first refers to the FOU ontology, develops some coding nodes according to the classes defined, and establishes an encoding system for semantic annotation and analysis of academic papers. During annotation, the encoding nodes are continuously expanded and adjusted according to the semantic content expressed in the academic paper, and the encoding system is updated. Finally, the hierarchical conceptual classes of PEO are determined, including 17 first-level classes such as background, research objectives, research significance, research content, methods, experiments, results, and conclusions, and 29 second-level classes obtained further subdividing the first-level class (see Table 1).

#### **3.2. Property Design in PEO**

Table 1 determines the classes and their hierarchical relationships, but it is inadequate. In order to fully express the paper's semantic units and their logical relationships, it is necessary to describe further the internal structure of these classes, where the structural information of these classes is the property of the class. This article designs the external properties and internal properties of the class. Among them, the former property is used to describe the relationship between the classes (semantic units in the paper), and the latter is the attribute information that describes itself.

### 3.2.1. External Property Design

In order to accurately describe the logical relationship between the above semantic units, this paper uses rhetorical relationships, argument relationships, chapter relationships, and knowledge element relations, plus custom relationships, to define a total of 18 logical relationships. They are the external property set in the ontology of academic papers (see Table 2).

**Table 1**  
Hierarchy concept class design of PEO.

First-Level Class	Second-Level Class	Co-occurrence Framework
Background	Background	SALT、 CoreSCs、 DEO、 FUO
Theme	Theme	SALT、 DoCO、 FUO
Problem	Problem	DEO
Research-Goal	Research-Goal	CoreSCs、 FUO
Research-Significance	Research-Significance	FUO
Research-Content	Research-Content	
Theoretical-Basis	Theoretical-Basis	
Definition	Definition	SciAnnotDoc、 FUO
Examples	Examples	
Data	Data	DEO
Conclusion	Conclusion	SALT、 CoreSCs、 DEO、 FUO
Future-Work	Future-Work	DEO、 FUO
Related-Research	Existing-Research	SciAnnotDoc、 FUO
	Research-Value	
	Research-Gap	
Method	Method-Paper	
	Method-Selection	FUO
	Method-Description	SciAnnotDoc、 CoreSCs、 FUO、 DEO
	Method-Advantage	
Experiment	Experiment-Environment	
	Experiment-Purpose	
	Experiment-Settings	
	Experiment-Content	CoreSCs
Result	Result-Description	CoreSCs、 DEO、 FUO
	Result-Description	CoreSCs、 DEO、 FUO
	Result-Metrics	
	Result-Evaluation	SALT、 DEO、 FUO
Discussion	Discussion-Recapitulation	DEO、 FUO
	Discussion-Limitation	FUO
	Discussion-Contribution	DEO、 FUO

### 3.2.2. Internal Property Design

Classes in PEO have some basic internal properties, such as information description, the article it belongs to, and the label information. In addition, in academic papers, authors cite and refer to the work of others. Therefore, some classes, e.g., background, existing research, and the paper's method, have

source information in the representation ontology of academic papers. Again, the author will hold a particular attitude or point of view. Therefore, some classes in the representation ontology of academic papers, such as research significance, research defects, results, and conclusions, often carry certain emotional information. Therefore, this paper defines five internal properties, with the specific contents listed in Table 3.

**Table 2**

External property set of PEO.

Property Name	Explanation	Refer(Source)
condition	A is the condition of B	RST
background	A is the background of B	RST
motivation	A is the motivation of B	customization
leads_to	A leads to B	SAO、 knowledge element ontology
review	A is the review of B	customization
introduces	A introduces B	customization
improves	A improves B	customization
resolves	A resolves B	customization
argues	A argues B	knowledge element ontology
produces	A produces B	SAO
supports	A supports B	SAO
not support	A does not support B	customization
summary	A is the summary of B	RST
purpose-behavior	Achieve A, B	discourse relationship
uses	A uses B	SAO
basis	A is the basis for B	customization
guides	A guides B	customization
elaboration	A is a elaboration of B	RST

**Table 3**

Internal property set of PEO.

Property Name	Property Value(Description)	Refer(Source)
Description	the content of the sentence(string)	customization
Article	the title of the article(string)	customization
Label	background, problem, method, result, conclusion etc.	customization
Tendency	positive, negative, neutral	SAO、 FOO
Source	other, own	SAO、 FOO

#### 4. Academic Paper Semantic Annotation Experiment

To evaluate PEO, this study utilizes the Nvivo data analysis tool [37] that exploits “deductive” coding. First, the encoding nodes are created according to the class in ontology, establishing the encoding system. Then, the sample data is encoded using the system, and finally, the annotation results are stored and analyzed. This study preprocesses the paper samples in PDF format and converts them into DOCX format used by Microsoft Word before annotating to remove diagrams, formulas, English abstracts, and references. The specific annotation process is illustrated in Figure 1.

## 4.1. Selection of Annotated Samples

Since articles in specific fields help analyze and compare results, based on previous work[20-21, 30] , we randomly selected 40 research papers published in 2017-2021 from Computer Science as annotated samples. This journal has a standard format, high quality, and reasonable length, and is more suitable for annotation experiments of academic papers.

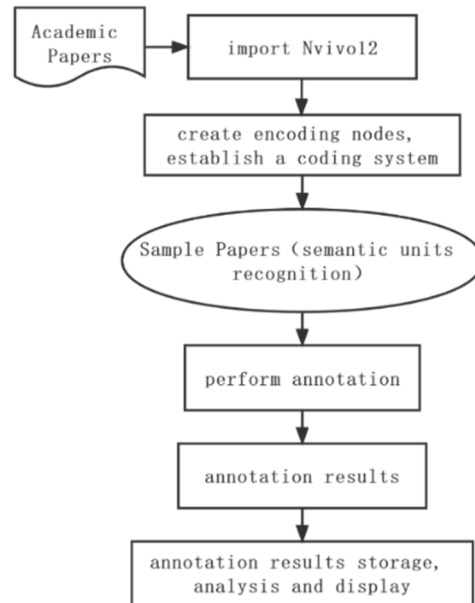


Figure 1: Academic paper semantic annotation process.

## 4.2. Annotation Experiment and Encoding Consistency Analysis

This research adopts the manual annotation strategy, which requires the annotators to judge and understand the content of the academic papers. Therefore, to ensure the reliability of the annotations, eight papers were randomly selected from a sample of 40 papers for consistency check, i.e., encoding consistency analysis, before starting the semantic annotation experiments. Specifically, first, the author annotated these eight papers, which were then annotated again by a person familiar with encoding conventions. Finally, the Kappa coefficient is calculated, an indicator used for consistency testing that can also measure the classification effectiveness [38]. The kappa coefficient is mostly between 0.6-1, presenting substantial consistency. After that, the author marked the remaining 32 papers and finally completed annotating the academic papers.

## 4.3. Annotation Result Analysis

Next, we statistically analyzed the annotating results. On the one hand, ontology coverage is used to evaluate the PEO coverage in all papers. On the other hand, text encoding coverage is used to assess the PEO coverage capabilities for individual papers, i.e., the ability of PEO to represent the semantic units of academic papers and their logical relationships verified from the above two aspects.

### 4.3.1. Ontology Coverage

Ontology coverage refers to the proportion of articles containing ontology categories in the total number of articles. Figure 2 illustrates the number of coding items of a single coding node. From the sample of 40 papers, different categories appear with different frequencies. Among them, nine categories such as “background”, “conclusion”, “outcome evaluation”, “method description”, and

“existing research” cover all academic papers, so these categories are regarded as common categories, illustrating the importance of this taxonomy. In addition, except for “theoretical basis”, “limitations”, “experimental environment”, “method selection”, and “research objectives”, the coverage rate of the remaining categories is more than 70%, which shows that most of the categories in PEO are representative.

### 4.3.2. Text Encoding Coverage

A node’s length proportion that encodes the content is important. By summing the encoding coverage of all categories in a single paper, the text encoding coverage of the entire paper is obtained to evaluate whether PEO can cover each academic paper. The statistical results are depicted in Figure 3, which reveals that the text encoding coverage is at least 75.33% and at most 92.57%, most of which falls in the 80.00% to 90.00% range. The average text encoding coverage rate of the 40 papers reached 84.64%. Therefore, the classes in PEO can express most of the academic paper content. To simplify the processing, some of the paper’s content has been appropriately deleted, e.g., figures, tables, and formulas, before annotating, while some content has not been annotated, e.g., keywords and titles at all levels. Therefore, the text encoding coverage is not statistically accurate, but it should be better than the results presented in the figure.

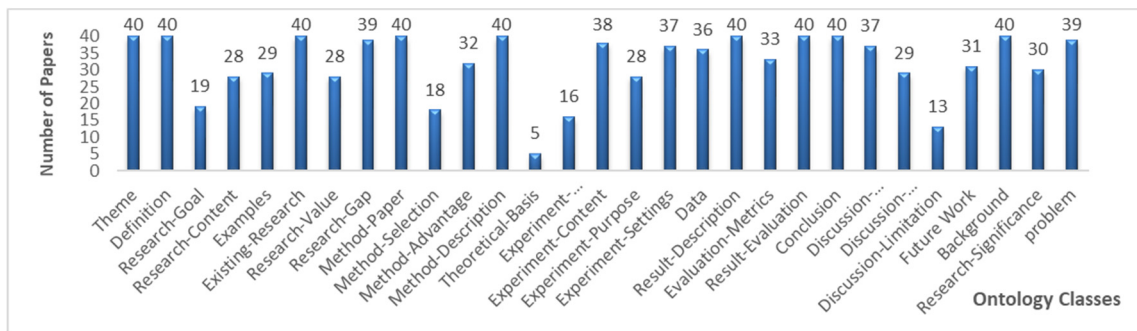


Figure 2: Statistical results of PEO ontology coverage.

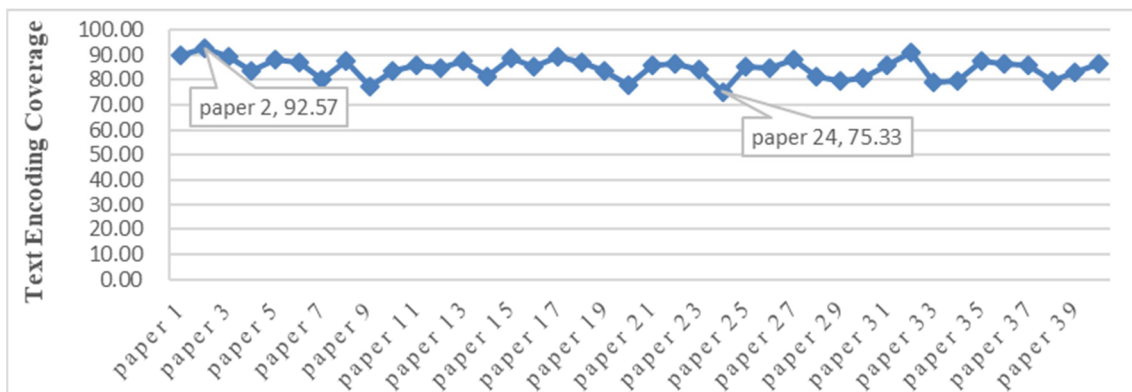


Figure 3: Statistical results of PEO text encoding coverage.

### 4.3.3. Comparison with Other Ontologies

In order to compare the representation ability of PEO, this paper uses the currently relatively mature Scientific Functional Unit Ontology (FUO) and Discourse Element Ontology (DEO) to annotate the same 40 sample papers. The corresponding results are reported in Table 4, highlighting that compared with FUO and DEO, the ontology coverage of the proposed PEO is 27.78% and 17.52% higher, respectively, and the text encoding coverage is 16.19% and 21.39% higher. These findings indicate that compared with existing ontologies, PEO has a stronger representation ability for the semantic units of academic papers.

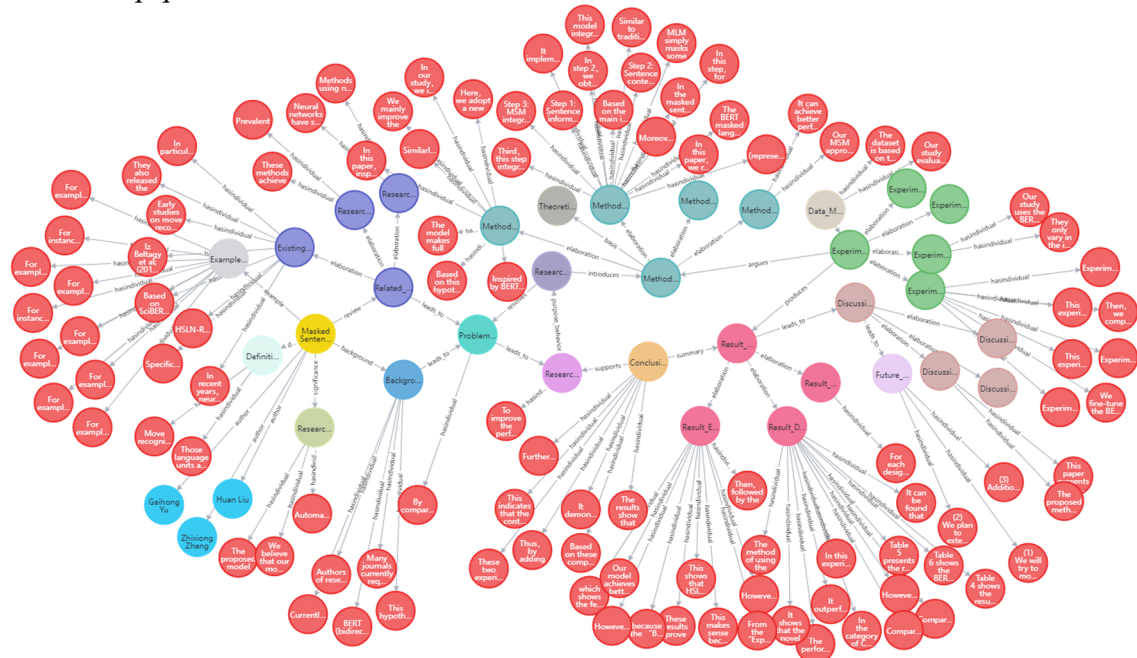


**Table 4**  
Annotation results based on different ontologies.

Ontology Name	Ontology Coverage (average)	Text Encoding Coverage (average)
FUO	51.96%	68.45%
DEO	62.22%	63.25%
PEO(this paper)	79.74%	84.64%

## 5. Knowledge Extraction and Storage of Academic Papers

Knowledge extraction and storage are important parts of a knowledge graph construction. Thus, first, this research uses the GATE (General Architecture for Text Engineering)[39] framework to semantically annotate two academic papers and obtain the documents in XML format. The titles of these two articles are “Moves Recognition in Abstract of Research Paper Based on Deep Learning” and “Masked Sentence Model Based on BERT for Move Recognition in Medical Scientific Abstracts”, from JDIL and JDIS, respectively. Then, the XML is parsed to obtain a series of instance data with semantic tags, and finally, the obtained instance data is mapped to the concepts of the ontology layer, and the Neo4j graph database is used for storage and visualization. Figure 4 visualizes the knowledge graph of the academic papers.



**Figure 4:** Example of PKG.

## 6. PKG Application Exploration

The knowledge graph constructed in this paper for the content of academic papers is only a prototype. In the future, the artificial processing link in the current process will be realized through intelligent means such as natural language processing and machine learning. At the same time, investigating knowledge graph fusion between multiple papers will also be considered. On this basis, the application of the knowledge graph is explored and realized in multiple directions.

Academic paper knowledge graphs and semantic technologies provide descriptions of the classification, attributes, and relationships of knowledge units in papers so that search engines can directly search for knowledge. For example, the user can directly query the “research objectives”, “background”, “research significance”, and “contribution” in a particular paper. As illustrated in Figure 5, this study presents a preliminary semantic retrieval example based on PKG. Realizing semantic retrieval can not only enable researchers to obtain information efficiently. At the same time, it can also

provide support for intelligent services such as intelligent question answering, decision support, and personalized recommendation.

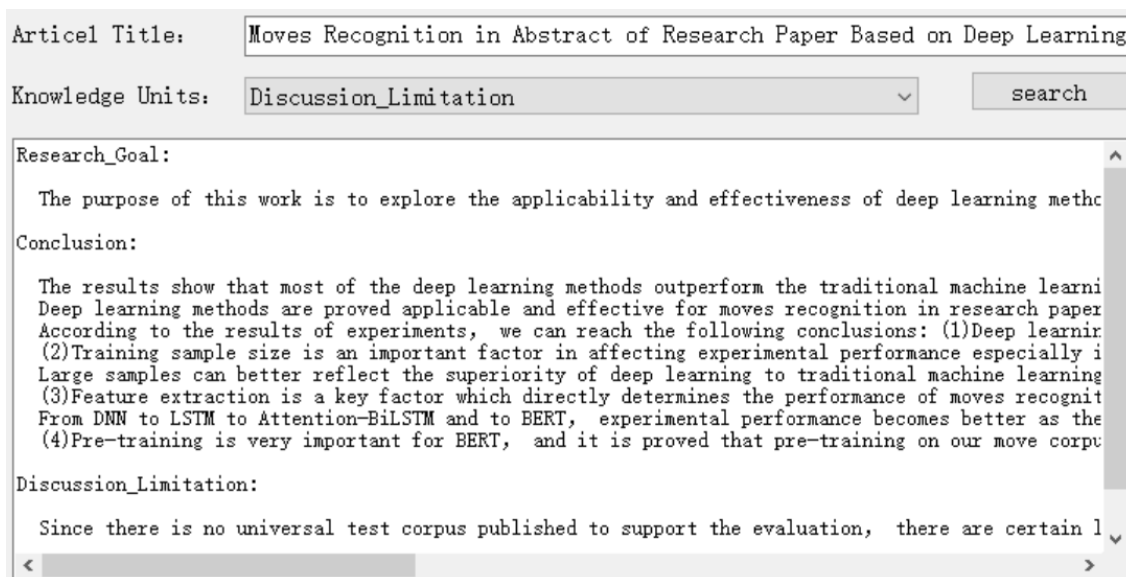


Figure 5: Application example of semantic retrieval PKG-based.

Automatically selecting or generating the corresponding responses according to some questions can improve the automation of information processing and resource acquisition efficiency and save human resources and costs. Based on the knowledge graph proposed in this paper, some intelligent questions answered in scientific research can be realized. For example: “Is there any literature mentioning that a certain method was used to solve a certain problem?”. As depicted in Figure 6, this study implements the above question and answer example based on PKG. The primary process of realizing this intelligent question answering is: first, parse the question sentence through advanced natural language processing technology, obtain the semantic information, and convert it into a query sentence in a structured form. Then, retrieve the relevant information from the knowledge graph and give relevant answers. In this way, researchers do not need to spend time and effort consulting literature but can quickly obtain relevant information from current research through the intelligent question-answering system to speed up scientific research.

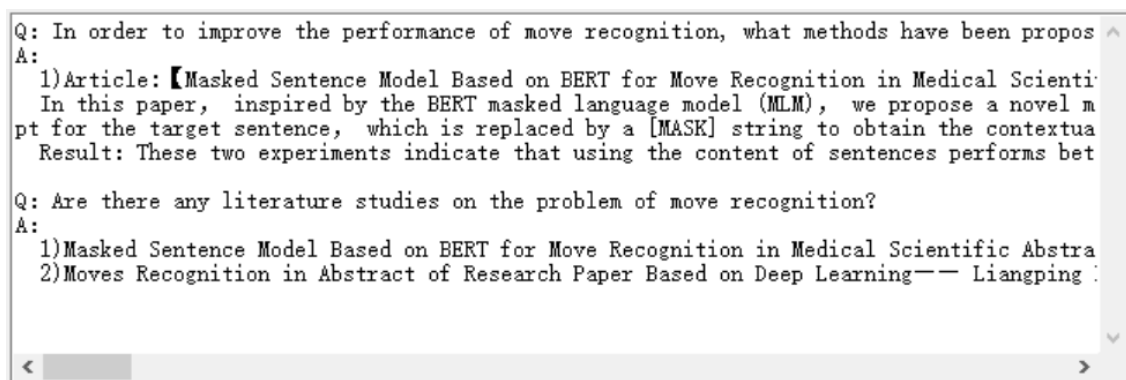


Figure 6: Application example of intelligent question answering PKG-based.

## 7. Conclusion and Outlook

Based on the knowledge units-theory, ontology, and knowledge graph theory and through the detailed analysis of the academic papers’ content, this study constructs an academic paper expression ontology (PEO), which solves existing research problems, such as too coarse modeling granularity and insufficient logical relationship representation ability. The semantic annotation experiment of academic

papers demonstrates that PEO ontology can comprehensively and deeply express the semantic units and their logical relationships in academic papers, verifying PEO's ontology ability to express academic papers. Second, we preliminarily construct the knowledge graph of academic papers based on PEO and through manual semantic analysis of the paper's content employing the GATE text annotation tool, XML parsing tool, and Neo4j graph database. Finally, semantic retrieval and intelligent question answering for academic knowledge are further realized based on PKG.

However, the current research still has some limitations. First, the PEO ontology only describes the text content semantically and does not consider other forms of content in the paper. Second, the knowledge graph construction process relies on manual analysis and processing. For the first problem, we design a particular semantic description model for the content outside the text format, combining the external features, internal features, charts, formulas, and other information ontologies or models of the paper to build a multimodal knowledge graph. This strategy covers academic knowledge in both breadth and depth. For the second problem, we employ natural language processing technology and machine learning technology for knowledge extraction and fusion to improve knowledge graphs' automatic construction. Furthermore, this strategy supports intelligent services such as semantic retrieval, intelligent question answering, intelligent recommendation, and automatic review generation for academic knowledge and information.

## 8. Acknowledgements

This work was supported by National Natural Science Foundation of China: the Intelligent Analysis and Optimization Method for Reflowable Documents(61672105). The English language was reviewed by EditSprings (<https://www.editsprings.cn>).

## 9. References

- [1] Ji, S. and Pan, S. and Cambria, E. and Marttinen, P. and Philip, S. (2021) A survey on knowledge graphs:Representation, acquisition, and applications, *IEEE Transactions on Neural Networks and Learning Systems*, 33, 494–514. <https://doi.org/10.1109/TNNLS.2021.3070843>
- [2] Zhang, F. and Liu, S. and Tang, J. and Dong, Y. and Yao, P. and Zhang, J. and *et al.* (2019) Oag: Toward linking large-scale heterogeneous entity graphs, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, USA, 2585–2595.
- [3] Tang, J. and Zhang, J. and Yao, L. and Li, J. and Zhang, L. and Su, Z. (2008) Arnetminer: extraction and mining of academic social networks, *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, Association for Computing Machinery, New York, USA, 990–998.
- [4] Sinha, A. and Shen, Z. and Song, Y. and Ma, H. and Eide, D. and Hsu, B. and Wang, K. (2015) An overview of microsoft academic service (mas) and applications, *Proceedings of the 24th international conference on world wide web*, Association for Computing Machinery, New York, USA, 243–246.
- [5] Bratsas, C. and Filippidis, P.M. and Karampatakis, S. and Ioannidis, L. (2018) Developing a scientific knowledge graph through conceptual linking of academic classifications, *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, IEEE, 113–118.
- [6] Auer, S. and Oelen, A. and Haris, M. and *et.al.* (2020) Improving access to scientific literature with knowledge graphs, *Bibliothek Forschung und Praxis*, 44, 516–529.
- [7] Fathalla, S. and Vahdati, S. and Auer, S. and Lange, C. (2017) Towards a knowledge graph representing research findings by semantifying survey articles, *International Conference on Theory and Practice of Digital Libraries*, Springer, Cham, 315–327.
- [8] Cao, S. and Zhao, B. (2021) The Construction and Application of Knowledge Graph for the Innovative Content of Academic Papers, *Journal of Modern Information*, 41, 28–37.
- [9] Roy, A. and Akrotirianakis, I. and Kannan, A.V. and Fradkin, D. and Canedo, A. and Koneripalli, K. and Kulahcioglu, T. (2020) Diag2graph: representing deep learning diagrams in research papers

- as knowledge graphs, 2020 IEEE International Conference on Image Processing (ICIP), IEEE, 2581–2585.
- [10] Wang, Y. (2020) Semantic Model for the Content of Scientific Literature, *Journal of Library and Information Science in Agriculture*, 32, 12–24.
- [11] Groza, T. and Handschuh, S. and Mollerer, K. and Decker, S. (2017) SALT-Semantically Annotated LATEX for Scientific Publications, *European Semantic Web Conference*, Springer-Verlag, Berlin, Heidelberg, 518–532.
- [12] de Waard, A. and Tel, G. (2006) The ABCDE Format Enabling Semantic Conference Proceedings, *Proceedings of 1st workshop: SemWiki2006-from wiki to semantics*, Budva, montenegro, 1–3.
- [13] Liakata, M. (2010) Zones of conceptualisation in scientific papers: a window to negative and speculative statements, *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, Association for Computational Linguistics, USA, 1–4.
- [14] WWW. (2011-12-22) Ontology of Rhetorical Blocks(ORB), [EB/OL]. <http://www.w3.org/TR/hcls-orb>
- [15] Di Iorio, A. and Vitali, F. and Peroni, S. (2013-07-16) The Pattern Ontology Describing documents by means of their structural components, [EB/OL]. <https://sparontologies.github.io/po/current/po.html>
- [16] De Ribaupierre, H. and Falquet, G. (2013) A user-centric model to semantically annotate and retrieve scientific documents, *Proceedings of the sixth international workshop on Exploiting semantic annotations in information retrieval*, Association for Computing Machinery, New York, USA, 21–24.
- [17] Shotton, D. and Peroni, S. (2015-07-03) The Pattern Ontology Describing documents by means of their structural components, [EB/OL]. <https://sparontologies.github.io/deo/current/deo.html>
- [18] Shotton, D. and Peroni, S. (2015-07-03) DoCO, the Document Components Ontology, [EB/OL]. <https://sparontologies.github.io/doco/current/doco.html>
- [19] Constantin, A. and Peroni, S. and Pettifer, S. and Shotton, D. and Vitali, F. (2016) The document components ontology (DoCO), *Semantic web*, 7, 167–181.
- [20] Qin, C.X. and Yang, Z.J. and Zhao, P.W. and Liu, J. (2018) The Knowledge Element Ontology Model of Scientific Literature for Knowledge Representation, *Library and Information Service*, 62, 94–103.
- [21] Wang, X.G. and Li, M.L. and Song, N.Y. (2018) Design and Application of Scientific Paper Functional Units Ontology, *Journal of Library Science in China*, 44, 73–88.
- [22] Zhang, L. and Kopak, R. and Freund, L. and Rasmussen, E. (2010) A taxonomy of functional units for information use of scholarly journal articles, *Proceedings of the American Society for Information Science and Technology*, 47, 1–10.
- [23] Sun, J.J. and Pei, L. and Jiang, T. (2018) Research on Semantic Annotation in Academic Literature, *Journal of the China Society for Scientific and Technical Information*, 37, 1077–1086.
- [24] Shotton, D. and Peroni, S. (2018-06-22) C4O, the Citation Counting and Context Characterization Ontology, [EB/OL]. <https://sparontologies.github.io/c4o/current/c4o.html>
- [25] Peroni, S. and Shotton, D. (2012) FaBiO and CiTO: ontologies for describing bibliographic resources and citations, *Journal of Web Semantics*, 17, 33–43.
- [26] Niu, H.L. and Ou, S.Y. (2020) Design and Application of a Semantic Annotation Framework for Scientific Articles, *Information studies: Theory & Application*, 43, 124.
- [27] Soldatova, L.N. and King, R.D. (2006) An ontology of scientific experiments, *Journal of the Royal Society Interface*, 3, 795–803.
- [28] Clark, T. and Ciccarese, P.N. and Goble, C.A. (2014) Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications, *Journal of biomedical semantics*, 5, 1–33.
- [29] Groth, P. and Gibson, A. and Velterop, J. (2010) The anatomy of a nanopublication, *Information Services & Use*, 30, 1–2.
- [30] Teufel, S. (1999) Argumentative zoning: Information extraction from scientific text, Ph. D. Dissertation. University of Edinburgh, Edinburgh, U.K.
- [31] Teufel, S. and Siddharthan, A. and Batchelor, C. (2009) Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics, *Proceedings of the 2009 conference on empirical methods in natural language processing*, 1493–1502.

- [32] Vitali, F. and Peroni, S. (2011-05-04) The argument model ontology, [EB/OL]. <https://sparontologies.github.io/amo/current/amo.html>
- [33] Wang, X.G. and Zhou, H.M. and Song, N.Y. (2020) Scientific Paper Argumentation Ontology and Annotation Experiment, *Journal of the China Society for Scientific and Technical Information*, 39, 885–895.
- [34] Qu, J.B. and Ou, S.Y. (2021) Semantic Modeling for Scientific Paper Argumentation Structure Driven By Semantic Publishing, *Journal of Modern Information*, 41, 48–59.
- [35] Mann, W.C. and Thompson, S.A. (2021) Rhetorical structure theory: Toward a functional theory of text organization, *Text-interdisciplinary Journal for the Study of Discourse*, 8, 243–281.
- [36] Chu, X.M. and Xi, X.F. and Jiang, F. and Xu, S. and Zhu, X.M. and Zhou, G.D. (2020) Macro Discourse Structure Representation Schema and Corpus Construction, *Journal of Software*, 31, 321-343.
- [37] Feng, D. (2020) *Qualitative Research Data Analysis Tool NVivo 12 Practical Tutorial*, Posts & Telecom press, Beijing.
- [38] Cohen, J. (1960) A coefficient of agreement for nominal scales, *Educational and psychological measurement*, 29, 37–46.
- [39] Cunningham, H. and Tablan, V. and Roberts, A. (2013) Getting more out of biomedical documents with GATE's full lifecycle open source text analytics, *PLoS computational biology*, 9, e1002854. <https://doi.org/10.1371/journal.pcbi.1002854>